AFRL-HE-AZ-TR-2006-0013

Transforming Training: A Perspective on the Need and Payoffs from Common Standards

Brian T. Schreiber

Lockheed Martin
6030 South Kent Street
Mesa AZ 85212

December 2006
Final Report for September 2003 to February 2006

Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Readiness Research Division

NOTICES

AFRL-HE-AZ-TR-2006-0013 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

Contract Number: F41624-97-D-5000 Task Order 34
Contractor: L-3 Communication-Link Simulation and Training

// Signed //

WINSTON BENNETT JR
Laboratory Task Monitor

// Signed //

HERBERT H. BELL
Technical Advisor

// Signed //

DANIEL R. WALKER, Colonel, USAF
Chief, Warfighter Readiness Research Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
30-12-2006

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
September 2003 to February 2006

4. TITLE AND SUBTITLE
Transforming Training: A Perspective on the Need and Payoffs from Common Standards

5a. CONTRACT NUMBER
F41624-97-D-5000 Task Order 34

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
63231F

5d. PROJECT NUMBER
4924

5e. TASK NUMBER AS

5f. WORK UNIT NUMBER
02

6. AUTHOR(S)
Brian T. Schreiber

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lockheed Martin
6030 South Kent Street
Mesa AZ 85212

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Readiness Research Division
6030 South Kent Street
Mesa AZ 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL; AFRL/HEA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-HE-AZ-TR-2006-0013

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Standards emerging today position the training research community on the eve of a scientific breakthrough. In the near future, the scientific community will likely benefit from the ability to routinely cross-compare training technologies and techniques from laboratory training study results, various operational training implementations, and possibly even live exercises. To achieve this capability, common standards must exist in the competencies to be assessed, the metrics used to evaluate those competencies, and the technology enablers to implement those assessments across training organizations. The current work aims (a) to discuss how the use of these common standards can afford this cross-comparison capability, (b) to provide a proof-of-concept study relying on only these standards, illustrating how this approach can be capitalized on at numerous training facilities, (c) to highlight where the common standards can be expanded, and (d) to provide some baseline distributed simulation within-simulator learning results. Thirty-five F-16 teams participated in week-long distributed mission operations (DMO) simulator training. The study was successfully conducted using the common standards with data captured on 31 teams. Minor issues were discovered in the technology enabling standards and recommendations are provided. By the end of the training week, F-16 teams increased weapons employment effectiveness and their kill ratios increased, while launching weapons at longer ranges and permitting fewer enemy strikers to reach their target. The results suggest assessing human performance across installations for cross-comparison of results is feasible, but some maturation of technology enabler standards is necessary to provide a routine, automatic, and robust inter-organizational cross-comparison capability.

15. SUBJECT TERMS
Air-to-air combat; Common standards; Distributed Mission Operations; DMO; Instructor ratings; Performance measurement; Standards; Training evaluation

16. SECURITY CLASSIFICATION OF:
Unclassified

a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED	22

19a. NAME OF RESPONSIBLE PERSON
Dr Winston Bennett Jr

19b. TELEPHONE NUMBER *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank.

TABLE OF CONTENTS

INTRODUCTION 1
Common Standard #1: Defining the Competencies.. 2
Common Standard #2: Metrics to Evaluate Competencies. 2
Common Standard #3: Technology Enablers. 4
Current Work 5
METHODS 5
Networked Simulation Facility 5
Participants.. 6
Training Research Syllabi 6
Training Research Week.. 6
Metrics 8
RESULTS 8
DISCUSSION... 10
REFERENCES 13
ACRONYMS.. 16

LIST OF FIGURES

Figure 1. An example standardized metric used to assess the Controls Intercept Geometry MEC skill. The friendly aircraft (in blue) ideally does not want the threat aircraft to penetrate any of the depicted zones with an aspect angle over 120 degrees (i.e., pointed at the friendly). 3

Figure 2 Example mirror-image point defense benchmark scenarios used for the pre- and post-test. 7

Acknowledgements

This research was performed under United States Air Force contract #F41624-97-D-5000, Task Order 34.

The author wishes to thank, in alphabetical order, the contributions of Dr. Winfred Arthur, Dr. Winston Bennett, Dr. Liz Gehr, Mary Johnson, Loren Keck, Les Pogue, Antoinette Portrey, Margaret Schurig, Dr. Bill Stock, and Eric Watz for their efforts in supporting the development of this report.

The views expressed herein are those of the author and do not necessarily reflect the views of Lockheed Martin or the United States Air Force.

TRANSFORMING TRAINING: A PERSPECTIVE ON THE NEED AND PAYOFFS FROM COMMON STANDARDS

INTRODUCTION

Contributions to the training research community occur at all levels, from basic research to integrating new advanced technology developments. From an application perspective, the use of common standards has the enormous potential to facilitate comparisons across laboratory and field studies. In contrast, application-oriented training research has typically been conducted using the idiosyncratic methods unique to particular institutions. This limits the scientific community's ability to provide guidance back to the warfighter when attempting to compare which inter-organizational methods and results do indeed yield the best value-added training. Scientifically powerful and unprecedented on a large scale, a common set of warfighter-valid methods and standards would allow for cross-comparison and the leverage of laboratory and field study results. This would permit quantifiable feedback to the warfighters as to which training techniques and technologies should be pursued. Standards emerging today position the training research community on the eve of this scientific breakthrough. In the near future, the scientific community is likely to benefit from this ability to routinely cross-compare training technologies and techniques from laboratory training study results, various operational training implementations, and possibly even live exercises. Retention and transfer-of-training studies could become routine.

To realize this scientific cross-comparison capability, common standards must exist in three primary areas, namely,

(1) defined skill competencies to be assessed,
(2) metrics to evaluate those skill competencies, and
(3) technology enablers to employ the assessment system across training sites.

Standards for defining warfighter competencies as well as standards for assessing warfighter performance against those competencies must first be established. Once warfighters have defined the core competency skill set and have devised metrics to measure performance on those skills, employing those competencies and metrics as standards for use across laboratory and field studies enables the cross-comparison of results for a given warfighter mission area. Of course, this requires that the technology mediums are in place to permit the implementation. This report discusses the importance of common standards in permitting for cross-comparison of results, reports a study demonstrating the proof-of-concept using only the common standards that would be necessary for study implementation at a number of sites, and advocates for technology enhancements that allow for expanding some of these standards to permit more comprehensive studies at any given site (Schreiber, Watz, & Bennett, 2003; Watz, Schreiber, Keck, McCall, & Bennett, 2003).

Common Standard #1: Defining the Competencies.

Determining common standards for the competencies which should be assessed is straightforward--the skills dictated by domain experts as necessary to perform their mission. Note that the training emphasis is on the Warfighter skill level/competency, not on the frequency with which a Warfighter practices an event (Chapman, Colegrove, & Greschke, in press). Once these mission essential skills have been identified and validated, these skills become the foundation that all training techniques and technologies should be evaluated against.

Laboratories and field sites purportedly training a given mission area or performing training research in a given mission area should use these standardized competencies as the basis for evaluation. The results from a process that reliably produces these competency skill sets provide the over-arching framework needed for defining a competency standard and for defining standardized metrics to assess those competencies.

Fortunately, a standardized process to define competencies already exists. Mission Essential Competencies (MECs) are “higher-order individual, team, and inter-team competencies that a fully prepared pilot, crew or flight requires for successful mission completion under adverse conditions and in a non-permissive environment” (Colegrove & Alliger, 2002). The MEC process uses only expert operational warfighter inputs as data, the results from which are both valid and reliable (Alliger, Beard, Bennett, Symons, & Colegrove, in press; Alliger, Garrity, See, McCall, & Tossell, 2004; Alliger, et al., 2003; Alliger, Colegrove, & Bennett, 2003; Colegrove & Alliger, 2002).

An example MEC air superiority skill is Controls Intercept Geometry (CIG); this skill entails managing inter-aircraft geometries such that the friendly aircraft minimizes vulnerabilities to the threats (while simultaneously being able to employ ordnance against the threat). In an air superiority mission, perfect performance, as defined by subject matter experts (SMEs), would result in desirable outcome metrics (e.g., no friendly mortalities, all threats killed) with flawless skill execution.

Common Standard #2: Metrics to Evaluate Competencies.

Obviously, once the competency skill set is defined, common standard metrics are needed to assess the skill competencies defined by the MECs across all training and training research installations. These metrics should exist at both the outcome and process (skill) level and be defined by SMEs as a direct subsequent step after the MEC process.

Consider air superiority: In a point defense mission, the overriding outcome objective is to deny enemy bomber aircraft within striking distance of the friendly point to be defended. The next most important outcome is to maximize the kill ratio—ideally killing all threats while all friendly aircraft survive. To consistently achieve these two standard high-level outcome objectives, warfighters must be proficient in the MEC skills. The air superiority skill CIG serves as an example. In the case of the CIG skill defined by SMEs, this would result in, ideally, never allowing a hostile fighter aircraft in any zone depicted in Figure 1 while that hostile is pointing at a friendly.

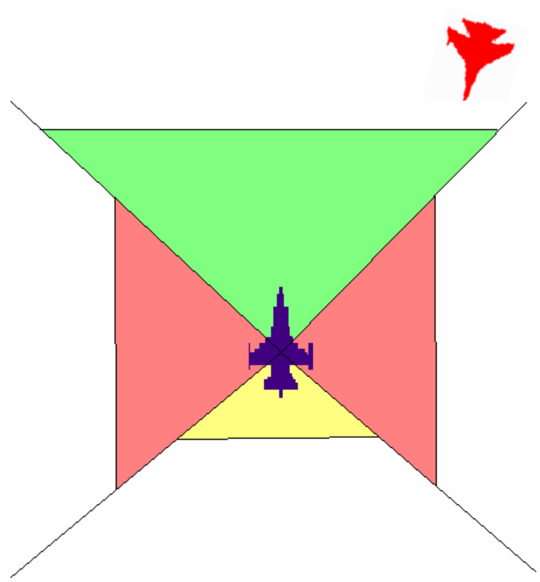

Figure 1. An example standardized metric used to assess the Controls Intercept Geometry MEC skill. The friendly aircraft (in blue) ideally does not want the threat aircraft to penetrate any of the depicted zones with an aspect angle over 120 degrees (i.e., pointed at the friendly).

Measuring the skill performance in addition to the outcomes will best reveal *how well* warfighters are performing at various skill competencies. For operational training, this will allow for standardized performance competency-based assessment across installations. Furthermore, the training research community is then better prepared not only to cross-compare at the outcome level, but also to identify which alternative training techniques and technologies are best for targeting which *skills*. With the ability to cross-compare at the outcome and skill level, the training community can both determine which training approaches yield the best mission outcomes and evaluate the specific skill improvement rates with the highest retention and transfer.

But, the "how" to assess the competencies can only be done if the technology is in place to support implementing those measurement standards across multiple laboratory and field training units. And, metrics such as CIG and others must be devised so that they can be captured in an automated fashion across a number of installations without developing new tools or customizations at those locations. "Automated performance measurement systems have been a required feature…but their application has been inconsistent and, in many cases, inadequate…" (Kelly, 1988, p.496).

Due to network protocol standards and recent performance measurement technology research, this inconsistent era may be ending, bringing us to the third and last major area required to enable routine scientific cross-comparison of results--the technology enablers in which to employ the standard metrics for assessing the MECs.

Common Standard #3: Technology Enablers.

In roughly the past decade, military training units and training research laboratories adopted networked simulators as a primary warfighter training method. In an effort to connect simulators allowing engagement in a virtual environment, engineers developed DIS, or Distributed Interactive Simulation (IEEE, 1995) and High Level Architecture (HLA) standards (e.g., Fischer, Case, & Bertin (Eds.), 2001). DIS or an agreed upon HLA Real-Time Interface (RTI) and Federated Object Model (FOM) requires all participating entities to supply standardized information across the computer network. Since these DIS and HLA network protocol standards are employed at most networked operational training and training research locations, a potential medium exists for incorporating standardized competency performance measurement, but it requires an assessment system to capitalize on this opportunity.

Schreiber, Watz, Bennett, & Portrey (2003) and Watz, Keck, & Schreiber (2004) discuss a Performance Effectiveness/Evaluation Tracking System (PETS) methodology exploiting the measurement distribution opportunity afforded by DIS and HLA so that assessment data at any DIS/HLA location can be theoretically captured. Simply stated, the PETS assessment system is another entity on the network adhering to the same DIS/HLA network protocol standards, not for the purpose of engaging other entities, but rather for reading network traffic to use as algorithm inputs to capture and record the metrics needed for assessing the MEC skills of various warfighters participating on that network. To provide an example of how DIS/HLA allows for standardized assessment of MECs, consider again the CIG skill. The warfighter's goal is to minimize the CIG time while achieving mission objectives. The CIG assessment rules defined by SMEs are (referring to Figure 1):

1. Identify hostile fighter aircraft and likely weapon load.
2. Determine hostile's aspect angle. If greater than 120 degrees (i.e., pointed at friendly), proceed with subsequent rules.
3. Determine hostile's quadrant (front, rear, side).
4. Determine hostile's altitude and range. (The altitude, range, and threat type dictate the critical ranges for each quadrant.)
5. Given the current altitude, range, and threat type, has the hostile penetrated within a critical range to friendly for that quadrant (Y/N)? If yes, increment time on CIG.

Therefore, to assess CIG for a given aircraft/warfighter, the following inputs are needed from any DIS/HLA network: Aircraft type, force affiliation (Red/Blue), position (latitude, longitude, and altitude), heading, and weapon type. The PETS system "listens" in real-time to the network traffic, "looking" for those inputs, then captures relevant inputs to identify the friendly and enemy fighters along with their altitudes and weapon types. The system continuously calculates (every 50 msec) all aspect angles and ranges between friendlies and threats. The end result is a simple determination by the PETS system whether or not any friendly has allowed a hostile to violate the abovementioned CIG rules, and the system increments a timer for each friendly that does so. Outcome metrics and additional process/skill metrics are captured in the same manner using the standardized DIS/HLA network traffic, which is rapidly becoming ubiquitous in the military training simulation community. MECs, metrics, and the technology enabling distribution system are the three instrumental standardized pieces necessary to allow for cross-comparison of laboratory and field training results.

Current Work

Since common standards for MECs, metrics, and the technology enablers of DIS/HLA and PETS now exist, a study utilizing only these standards is all that is necessary to demonstrate a proof-of-concept for permitting routine scientific assessments. The results of the study, especially the lessons learned, also serve to highlight where the common standards need to mature further. The search was short for finding a suitable study that could not only serve this proof-of-concept, but also contribute to the scientific body of research.

Defying good business practice, DIS/HLA networked simulations are rapidly becoming the warfighter training medium of choice without the backing of literature supplying objective, quantifiable in-simulator performance improvements--a disturbing trend which is far from new (Waag, 1991). Therefore, a fundamental within-simulator training effectiveness study documenting the learning taking place within a distributed simulation environment would perfectly satisfy current study requirements.

METHODS

Networked Simulation Facility

The Distributed Mission Operations (DMO) training research facility at the Warfighter Readiness Research Division in Mesa, AZ, provided the distributed simulation environment used for the present study. Four high-fidelity F-16 simulators and one high-fidelity Airborne Warning and Control System (AWACS) were used in conjunction with a computer-generated threat system and an instructor operator station (IOS). Similar to many distributed simulation training environments, all entities interoperated according to common DIS standards.

The high-fidelity F-16 simulators were Block 30 with a 360 degree out-the-window visual display. The F-16 display systems used either SGI® Onyx2 Reality Monster Visualization supercomputers or pC-Novas (v2.0) running Aechelon runtime software. The visual system used high resolution photo-realistic databases of the Sonoran desert overlaid on terrain elevation data of the region. The hardware in the cockpits was identical to that found in the actual F-16, as was the software (Software Capabilities Upgrade [SCU] version 4). Depending on the type of mission to be flown, F-16 weapon load-outs for missions consisted of differing combinations of the gun, the Air Intercept Missile (AIM-9), the Advanced Medium Range Air-to-Air Missile (AMRAAM), and/or the Mk-82 and Mk-84 general purpose bombs. A high-fidelity AWACS sensor simulation was also used to provide a more realistic environment. The high-fidelity AWACS station was a Solipsys MSCT V. 3.9 networked to the Solipsys TDF V. 2.7.3.

The computer-generated threat system used was the Automated Threat Engagement System (ATES). ATES is a real-time threat generation system for use on a standard DIS network. The ATES system uses aerodynamic modeling, atmospheric models, radar models, infrared (IR) models, and data parameter tables for thrust, drag, lift, etc. For the current work, threat air models were the MiG-29, MiG-27/23, and Su-27 loaded with the AA-8, AA-10a and AA-10c air-to-air missiles. Ground threats included the SA-2, SA-6, and SA-8, and antiaircraft artillery (AAA). Threat aircraft followed maneuvers and/or scripted flight paths and reacted to friendly maneuvers and weapons.

Participants

Operational F-16 pilots and AWACS controllers routinely visit the Warfighter Readiness Research Division in Mesa, AZ for participation in various training research studies. For the current work, 35 operational F-16 teams (four fighter pilots fly as part of a four-ship team) who participated in five-day training research between January 2002 and May 2003 were used. The mean number of hours flown in the F-16 was 964 (range 448 to 2088).

Training Research Syllabi

During the data collection period, pilots flying the F-16 simulators “flew” one of four very similar syllabi--each syllabus consisted of nine sessions, beginning with session one on Monday morning and ending with session nine on Friday morning. There were two sessions each day of the five-day training week, except for Friday when the participants had only one session. Each session entailed a one-hour briefing, an hour of flying, and an hour and a half debriefing.

The syllabi scenarios could be either offensive or defensive, but all consisted of four F-16s versus X number of threats. Scenarios were designed with trigger events and situations to specifically train MEC skills. These syllabi were developed with traditional methods using full mission rehearsal scenarios across a spectrum of probable air-to-air missions and threats while increasing the complexity of the missions as the training research week progressed.

Training Research Week

Each syllabus began with a familiarization session (session one) to orient pilots to DMO simulator environment specifics, such as visual identification (ID) characteristics and any switchology differences due to F-16 block number or F-16 mission software. The pilots required very little familiarity training, since the high-fidelity simulator layout closely resembled the actual aircraft and since all the declarative and procedural knowledge to be operationally qualified to fly the F-16 had been learned by participants before arriving. Therefore, after the familiarity session, performance increases observed throughout the course of the subsequent sessions were the result of learning how and when to best employ the skills they had been taught during their Air Force career.

Session two (after the familiarization period) began with benchmarks (i.e., a “pre-test”) used to measure pre-training performance. The benchmarks consisted of flying three point defense engagements (see Figure 2). All benchmark point defense scenarios pitted the four participant F-16s against eight threats (six hostiles and two strikers); all benchmarks were designed to be equally complex according to the absolute complexity scoring scheme outlined by Denning, Bennett, and Crane (2002).

Five-point defense benchmark scenarios were developed, and the complexity analysis revealed that all benchmarks were indeed equally complex. Unbeknown to the pilots, for the Friday benchmarks, participants (in the same flight/cockpit assignment) flew the mirror-image of the three benchmarks that were flown on Monday.

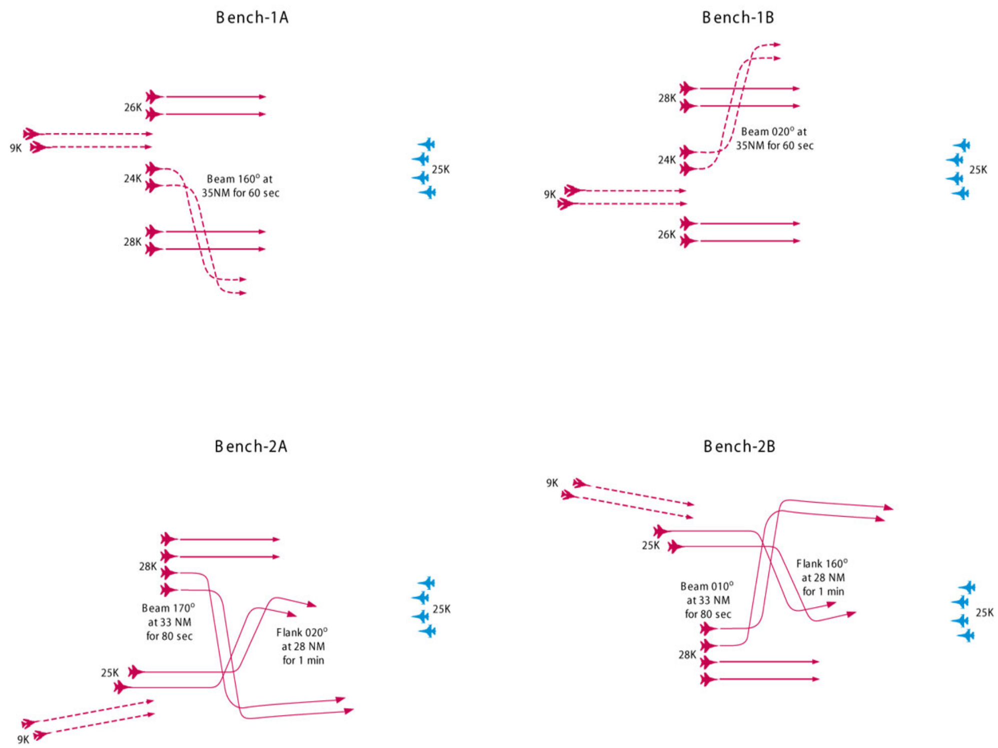

Figure 2 Example mirror-image point defense benchmark scenarios used for the pre- and post-test.

The participants' overriding goal for the point defense benchmark scenario was to prevent the enemy striker/bombers from reaching the base – success being striker denial or kill. The benchmark scenarios were selected for examination in the present study as pre- and post-test assessments because:

(1) all the benchmark engagements have equivalent levels of complexity,

(2) three benchmark scenarios occur at the beginning and the end of the week-long DMT syllabus,

(3) the same pilots perform the benchmark scenarios in the same team positions at the beginning and the end of the week, and

(4) the benchmarks were flown under real-time kill removal and strict data collection rules.

The MEC-based building-block training began immediately after the benchmarks and continued through the course of the week. Participating teams were exposed to four to eight full engagements per session, with each engagement generally concluding with a logical end such as "Bingo" (nearly out of fuel), all threats killed, or multiple friendly losses. While these training sessions emphasized Defensive Counter Air (DCA) scenarios, pilots also flew Offensive Counter Air (OCA) and air-to-ground missions. All engagements were flown versus simulation of actual threat aircraft, air-to-air ordnance, and surface-to-air ordnance. These 30+ engagements between benchmarks provided a very rich environment for air-to-air training and were the equivalent of flying more than ten friendly four-ship missions, with each mission opposed by 8-16 dissimilar adversary aircraft. The training sessions also provided real-time enemy kills and real-time friendly losses. The building block training sessions progressed in complexity by increasing the number of threat aircraft, the type of threat aircraft, the threat aircraft reactivity/maneuver, and/or an increase in the vulnerability time.

Metrics

A primary goal of the current work was to demonstrate a proof-of-concept study relying on only common standards for obtaining the data, thereby illustrating that the groundwork for leverage and cross-comparison of laboratory and field studies is possible. As such, the metric of greatest interest is the success or failure of the MECs, metrics, and technology enablers (DIS and PETS) for conducting a distributed simulation study.

For the secondary goal of providing baseline within-simulator effectiveness data, all metrics were therefore captured using only standards methodologies discussed. DIS and PETS provided the standardized technology enablers for capturing the MEC-based outcome and skill metrics.

For outcome metrics, enemy strikers reaching target, enemy kills, friendly mortalities, and percentage of threat and friendly shots resulting in a kill were processed and recorded using PETS and only the information available from the DIS network.

For skill metrics, the MEC skill CIG and one indicator of the MEC weapons employment skill--weapons launch range--were recorded in the same manner and are reported in the current work. To report substantially more process metrics for more skills, a more comprehensive limited distribution technical report documenting within-simulator learning is currently in preparation. Only high-level descriptive statistics, in terms of percentage change, are reported here.

RESULTS

There were four major result areas of interest. Each of the first three revolved around the success or failure of using the three pivotal common standard areas previously discussed, MECs, metrics, and the underlying technology enabler system (DIS/HLA and PETS). Fulfilling our secondary objective, the fourth result area was to report initial within-simulator training results that could serve as a baseline for future cross-comparisons when evaluating alternative training techniques and technologies. Results in each of the four areas are discussed in turn.

The MEC process for identifying critical skills in air superiority predated the study here and, save some delays in obtaining operational personnel for data collection, posed no issues. The MEC process produced 37 skills required for successful air superiority in actual combat under adverse conditions, thereby providing the common standard defined skill set for warfighter air superiority metric development.

Devising standard metrics for each of the MEC skills required a diverse set of solutions. Common standards for MEC air superiority outcome metrics (e.g., strikers on target, kill ratios) were quickly and easily identified. Skill metrics, however, were much more varied, some simple, others more complicated. Some MEC air superiority skills, such as "weapons employment," were rapidly (and unanimously) identified by SMEs. The result was to capture a number of different data points at weapon launch and weapon detonate (e.g., launch range, launch altitude, launch airspeed, distance between entities at weapon detonate, etc.).

Other MEC skills, such as the CIG metric already described, required decomposing SME heuristics into rule sets suitable for translation into programmable code for the PETS system to capture off of a standard DIS/HLA network. A number of the MEC air superiority skills could not be converted into objective metrics for capture by the PETS assessment technology (e.g., communication). For these MEC air superiority skills, separate subjective assessment tools were used.

In the end, all MEC air superiority outcome metrics were successfully captured, some skill metrics were successfully captured objectively and off-line subjectively, and the remaining skill metrics are still in development with a new subjective assessment system that ultimately is designed for incorporation into the PETS technology (MacMillan, Entin, & Morley, in press).

The DIS and PETS enabling technologies used to capture the objective metrics, worked generally as expected. The DIS environment met objectives and most expectations, allowing all entities to interoperate routinely and successfully on over 1,000 simulated engagements with only a few notable DIS issues. As such, data was successfully captured according to research protocol for 31 teams. "According to research protocol" was a logistical and control necessity added due to a limitation in the DIS network protocol standard. No standards within the DIS community exist to regulate the human operators of the DIS distributed simulation environments. That is, under DIS and HLA common protocol standards, console operators are free to act in "God-like" manners that would largely invalidate any conclusions drawn from metrics obtained. Examples include using "shields," regenerating killed entities, reloading fuel without an aircraft visiting a tanker, etc. These common standards limitations were addressed early in this study by writing specific "research protocols" to be adhered by all operators. Additionally, early in the current research it was discovered that DIS protocol standards only mandate a limited and narrowly focused set of data to be shared among entities on the network (i.e., mainly positional and attributional data), thereby limiting the potential for the PETS system to collect the necessary data for some objective metrics. Furthermore, because DIS operates on a broadcasting protocol (i.e., no recipient confirmation required), some standard data packets could go missing, resulting in that data never being processed by the PETS technology.

Finally, though tangentially germane to the current work, it is of interest to note that during other, non-related large-scale research exercises, additional DIS issues would occasionally surface, such as bandwidth and DIS version control.

In summary, most engineering issues for this study (single site, less than twenty entities) occurred not so much with DIS, but with a single simulator or the threat generation system (i.e., problem resided within the simulator system itself, not the DIS network protocol).

Contradicting pre-study suppositions, the PETS enabling technology used to collect the metrics worked quite well for skill metrics, but outcome metrics were much more complicated to capture. Much of this difficulty could be attributed to occasional inconsistencies in the DIS network (e.g., missing network data traffic described above. Other difficulties required updating the PETS technology to capture unique events impacting outcome metrics (e.g., correctly registering a kill when one aircraft chases another into a mountain without shooting it). These challenges were overcome with updated code and all objective outcome and skill metrics reported here were collected automatically and successfully.

Baseline within-simulator training effectiveness study results revealed that all metrics for the 31 teams showed improvements in the expected direction. Compared to the Monday benchmarks (session two), performance observed on the Friday benchmarks (session nine) showed 69% fewer F-16 mortalities, 61% fewer enemy bombers reaching base, 25% more threats killed, 10% longer range at launch of missile, 69% improved performance (less time) on the MEC CIG skill metric, 55% fewer threat shots resulting in a kill, and 7% more F-16 shots resulting in a kill (Gehr, Schreiber, & Bennett, 2004).

DISCUSSION

The successful distributed simulation study reported here represents a training transformational capability to automatically capture objective human performance data from a DMO environment. Relying only upon the MEC, metric, and enabling technology standards, this study illustrates the potential capabilities for the training community in the future. Though the current work reported only one study at one location, the study was conducted by relying on standards that should theoretically be easy to apply at another DMO location. Indeed, based upon the promising feasibilities, efforts are currently underway to enable and test these assessment capabilities at a sample of other DMO sites (e.g., Shaw Air Force Base; Bills & Devol, 2003). A demonstration for collecting the same standardized metric data is also planned for a live fly event at Nellis Air Force Base by the end of 2005.

With a capability to standardize assessing skill competencies across field site and laboratory installations, the operational community would be able to, at any time and at any place, theoretically assess a warfighter on his/her skill and carry those results forward longitudinally and across installations. Furthermore, the scientific community would be afforded the ability to cross-compare study results evaluating alternative training techniques or technologies and do so quantitatively, thereby revealing the best value added training approaches.

The processes for MEC development and metric development, though SME-intensive, did not pose any significant issues to conducting this study. This makes intuitive sense, as both those common standards are processes which result in information standards to be used. Of course, a proof-of-concept study such as this, while successful, was not without complications. The success or failure for researchers to take those information standards and convert them into application and concrete, valid assessments for this (or any future study) hinges upon the technology enablers. Therefore, it was expected that the majority of issues encountered would exist with the DIS and PETS enabling technologies.

The DIS/HLA standards are rather limited (Lacy & Tuttle, 1994) and should expand. In DIS and commonly used HLA FOMs, the typical data packet passed between interoperating entities on the network contains roughly 13 variables. The entity state packet, for example, contains primarily attribution and positional information (e.g., Su-27, latitude, longitude, altitude) updated at a given frequency. The variables within these data packets are the only sources of information for which standardized assessment methodologies such as PETS can use as network inputs for calculating performance metrics. Additional inputs can, however, be taken from non-network sources such as configuration tables, as the CIG assessment algorithm does for its required quadrant ranges.

Since network traffic is minimal, the pool of input variables for assessments is limited. Consider the CIG metric: All the data required for computing that metric is available per common DIS/HLA standards or simple configuration tables, except for the threat's weapons load. Therefore, a custom modification was performed within DIS standards to allow for capturing and assessing the CIG metric accurately--obviously not the desired long-term solution. The other undesirable option would have been to operate under an assumed weapons load given the type of threat—the threat type being known from network traffic. But this approach introduces errors when those assumptions are not true. The current limited network data traffic exists primarily out of meeting only basic interoperating needs and bandwidth limitations. Given standardized assessment requirements and time to allow technology enhancements to increase bandwidth for DIS/HLA environments, more MEC skills could be assessed using standardized metrics and standardized technology enablers such as PETS. If standards do not expand, only outcome metrics and a limited set of MEC skill metrics can be automatically and objectively obtained via any standard DIS/HLA DMO network.

Perhaps less obvious for cross-comparing and leveraging results between organizations, standards for administering distributed simulation events should exist. Conveyed more clearly by way of example, consider the "regeneration" capability. Using regeneration from the IOS during an unfolding scenario impedes attempts to automatically collect outcome metrics (e.g., kill ratios). Even with extensive code to accurately collect this information in spite of IOS operator "God-like" actions, the data are rendered almost useless for interpreting and drawing conclusions about the training. Other IOS functionalities carry similar assessment pitfalls, such as shields, freezing, or relocating entities. These approaches may very well be desirable as part of a training technique or strategy, but for measurement points (i.e., benchmarks) to formally assess performance, the "realism" approach is best—using kill removals, no mid-air weapons reloading, no refueling in flight unless done so via tanker, etc. In addition to standardizing measurement points, this approach also provides stronger conclusions to be drawn about the value of the training and allows for more direct comparisons to range exercises.

The demonstrated performance improvement results suggest that significant learning took place in the DMO environment. These results provide strong evidence for reaffirming some DMO training effectiveness subjective data studies (Bennett, Schreiber, & Andrews, 2002; Crane, Robbins, & Bennett, 2000; Krusmark, Schreiber, & Bennett, 2004; Waag, Houck, Greschke, & Raspotnik 1995), but the conclusions here are taken further by quantifying the magnitude of in-simulator learning improvement. These objective results show the F-16 teams were not simply sacrificing performance in one area to improve performance in another area, but rather that they were improving in both offensive and defensive skills. By the end of the training week, F-16 teams performed the MEC CIG skill more effectively, they increased weapons employment effectiveness, and their kill ratios increased--all while launching weapons at longer ranges and permitting fewer enemy strikers to reach their target.

In addition to learning other critical skills, it is postulated that the F-16 pilots learned where and when to best position their weapons systems in specific inter-aircraft geometries such that they could effectively employ their radar missiles, but simultaneously avoid vulnerable exposure to the threats' weapons engagement zones. The current study can be used as a baseline DMO training effectiveness study which other laboratory studies or operational DMO sites can then compare against when evaluating alternative training approaches.

Direct comparisons to range exercises provide the final, long-term objective—using the same standards for assessing performance and cross-comparing results from training research laboratories, operational training locations, and range exercises. For example, at the Nellis Range, much of the data for the aircraft participating in live-fly exercises is passed in a similar manner to current DIS/HLA network protocol standards. If the standards already discussed can be employed not only at the DMO simulation facilities, but also at the live exercise ranges, the scientific potential for discovering the best uses of DMO cannot be overemphasized. Objective, in-simulator learning assessments could become routine and thus any systematic change within or between similar DMO environments could then be objectively assessed. Furthermore, straightforward transfer of training assessments from the DMO environment to the range becomes possible.

It appears that this cross-comparison era is dawning. As mentioned, DIS and HLA are already commonly accepted network protocol standards. The United States Air Force's Air Combat Command (ACC) has called for MECs to be developed for all major Air Force weapons systems (over 15 of which are either in process or completed), and the metrics and PETS assessment methodology have been identified as ACC's potential solution for an Air Force-wide MEC competency-based assessment system. The backing of these communities solidifies these necessary core common standard areas as standards likely to be implemented across a great number of military training and training research institutions, creating a transformation for training research, warfighter competency-based training, and evaluating alternative training techniques and technologies.

REFERENCES

Alliger, G.M., Beard, R., Bennett, W. Jr, Symons, S. R., & Colegrove, C.M. (in press). A psychometric examination of Mission Essential Competency (MEC) measures used in Air Force Distributed Mission Operations training needs analysis. *Military Psychology*, Special Edition.

Alliger, G. M., Garrity, M. J., See, K., McCall, J. M., & Tossell C. (2004). Advanced training for commanders: A competency-based approach to training requirements definition for the JFACC. In *2004 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Alliger, G. M., Garrity, M. J., McCall, J. M., Beer, L., & Rodriguez, D. (2003, April). *Competency-based definition of work and performance for command and control.* Paper presented at the Thirteenth International Occupational Analyst Workshop. San Antonio, TX.

Alliger, G., M., McCall, J. M., Garrity, M. J., Morley, R. M., Beer, L., & Rodriguez, D. (2003). Mission Essential Competencies for the AOC: A basis for training needs analysis and performance improvement. In *2003 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Alliger, G. M., Colegrove, C. M., & Bennett, W. Jr. (2003, April)**.** *Mission Essential Competencies: New method for defining operational readiness and training requirements.* Paper presented at the Thirteenth International Occupational Analyst Workshop. San Antonio, TX.

Bennett, W. Jr., Schreiber, B. T., & Andrews, D. H. (2002). Developing competency-based methods for near-real-time air combat problem solving assessment. *Computers in Human Behavior,* 18, 6. 773-782.

Bills, C.G. & Devol, D. (2003). How F-16 MTC team achieved operational training capability. In *2003 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Chapman. R., Colegrove, C. M., & Greschke, D. A. (in press). Training transformation in theory and practice. *Military Psychology*, Special Edition.

Colegrove, C. M. & Alliger, G. M. (2002, April). *Mission Essential Competencies: Defining combat mission readiness in a novel way*. Paper presented at the NATO RTO Studies, Analysis and Simulation (SAS) Panel Symposium. Brussels, Belgium.

Crane, P. M., Robbins, R., Bennett, W. Jr. (2000). Using Distributed Mission Training to augment flight lead upgrade training. In *2000 Interservice/Industry Training, Simulation* *and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association. (AFRL-HE-AZ-TR-2000-0111, ADA394919). Mesa, AZ: Air Force Research Laboratory, Warfighter Readiness Research Division.

Denning, T., Bennett, W. Jr., Crane, P. M., & Bell, J. A. (2002). Mission characteristics and cComplexity in Distributed Mission Training. In *2002 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Denning, T., Bennett, W. Jr., & Crane, P. M. (2002). Mission complexity scoring in Distributed Mission Training. In *2002 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Fischer, J., Case, B., & Bertin, R. (Eds) (2001). *Guidance, rationale, and interoperability manual for the Real-time Platform Reference Federation Object Model (RPR FOM). Version 2.0D9v2.* Orlando, FL: Simulation Interoperability Standards Organization.

Gehr, S. E., Schreiber, B.T., & Bennett, W. Jr. (2004). Within-simulator training effectiveness evaluation. In *2004Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Institute of Electrical and Electronics Engineers. (1995). *IEEE Standard for Distributed Interactive Simulation – Application Protocols* (IEEE Std 1278.1-1995). New York: IEEE Computer Society.

Kelly, M. J. (1988). Performance measurement during simulated air-to-air combat. *Human Factors, 30(4), 495-506.*

Krusmark, M. A., Schreiber, B. T., & Bennett, W. Jr. (2004). *The effectiveness of a traditional gradesheet for measuring air combat team performance in simulated Distributed Mission Operations* (AFRL-HE-AZ-TR-2004-0090, AD A428119). Mesa, AZ: Air Force Research Laboratory, Warfighter Readiness Research Division.

Lacy, L. & Tuttle, C. (1994). Use of the set data PDU and the event report PDU to support data analysis and reporting. In *Proceedings of the 10th Workshop on Standards for the Interoperability of Defense Simulations,* March 14-18, 1994.

MacMillan, J., Entin, E. B., & Morley, R. M. (in press). Measuring team performance in complex and dynamic military environments: The SPOTLITE method. *Military Psychology*, Special Edition,

Schreiber, B. T., Watz, E. A., & Bennett, W. Jr. (2003). Objective human performance measurement in a distributed environment: Tomorrow's needs. In *2003 Interservice/ Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Schreiber, B. T., Watz, E. A., & Bennett, W. Jr. (2003). Development of a Distributed Mission Training automated performance tracking system. In *Proceedings of the Behavioral Representations in Modeling and Simulation (BRIMS) Conference*. Scottsdale, AZ.

Waag, W. L. (1991). The value of air combat simulation: strong opinions but little evidence. In *Training transfer - Can we trust flight simulation? Proceedings of the Royal Aeronautical Society Conference on Flight Simulation and Training.* London: Royal Aeronautical Society.

Waag, W. L., Houck, M. R., Greschke, D.A., & Raspotnik, W. B. (1995). Use of multiship simulation as a tool for measuring and training situation awareness. In AGARD Conference Proceedings 575 *Situation Awareness: Limitations and Enhancement in the Aviation Environment* (AGARD-CP-575). Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research & Development.

Watz, E. A, Keck, L., & Schreiber, B. T. (2004). Using PETS software to capture complex objective measurement data from Distributed Mission Operations (DMO) environments. In *Proceedings of the 2004 Spring Simulation Interoperability Workshop*. Arlington, VA. Paper available as 04S-SIW-143.

Watz, E. A., Schreiber, B. T., Keck, L., McCall, J. M., & Bennett, W. Jr. (2003). Performance measurement challenges in Distributed Mission Operations environments. [Electronic Version] *Simulation Technology Newsletter.* 03F-SIW-022. March 16, 2004. retrieved 7/26/04 from http://www.sisostds.org/webletter/siso/Iss_102/index.htm 2003 Fall SIWzie Award winner - Fall Simulation Interoperability Workshop

ACRONYMS

Antiaircraft Artillery	AAA
Air Combat Command	ACC
Air Intercept Missile	AIM-9
Advanced Medium Range Air-to-Air Missile	AMRAAM
Automated Threat Engagement System	ATES
Airborne Warning and Control System	AWACS
Controls Intercept Geometry	CIG
Defensive Counter Air scenarios	DCA
Distributed Interactive Simulation	DIS
Distributed Mission Operations	DMO
Federated Object Model	FOM
High Level Architecture	HLA
Identification	ID
Infrared	IR
Institute of Electrical and Electronics Engineers	IEEE
Instructor Operator Station	IOS
Mission Essential Competencies	MECs
Offensive Counter Air	OCA
Performance Effectiveness/Evaluation Tracking System	PETS
Real-Time Interface	RTI
Software Capabilities Upgrade	SCU